NDMI113 – Extremal Combinatorics

Mykhaylo Tyomkyn

Lecture 5 - Exact and modular intersections.

In the previous lecture we considered set systems $\mathcal{F}$ where every two distinct members intersect in at least t elements. What if instead we demand that every intersection is *exactly* of size t: how large can $\mathcal{F}$ be? For $t = 0$ it is easy to see that the answer is $n+1$, attained by $\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \ldots, \{n\}\}$. For $t = 1$ one plausible candidate to be the largest is $\mathcal{F} = \{1\}, \{1,2\}, \{1,3\}, \ldots, \{1,n\}\}$, with $|\mathcal{F}| = n$. This turns out to be best possible for any $t \geq 1$.

Theorem 1 (Fisher's Inequality) *Let $t \geq 1$ and $\mathcal{F} \subseteq 2^{[n]}$ with $|A \cap B| = t$ for all $A \neq B \in \mathcal{F}$. Then $|\mathcal{F}| \leq n$.*

Proof If for any $A \in \mathcal{F}$ we have $|A| = t$, then $A \subseteq B$ for all $B \in \mathcal{F}$. Since $|B \cap B'| = t$ for all $B, B' \in \mathcal{F}$, the sets in $\mathcal{F}_A = \{B \setminus A \mid B \in \mathcal{F}\}$ are disjoint. Hence, $|\mathcal{F}| \leq n$.

So, we may assume that $|A| > t$ for all $A \in \mathcal{F}$. For each $A \in \mathcal{F}$ consider its characteristic vector χ_A over $\mathbb{Q}$, defined as $\chi_A = (\chi_A(1), \ldots \chi_A(n)) \in \mathbb{Q}^{[n]}$, where

$$\chi_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases}$$

For instance, if $\mathcal{F} = 2^{[5]}$ and $A = \{1,2,5\}$, then $\chi_A = (1,1,0,0,1)$.

We claim that the set of vectors $\{\chi_A \mid A \in \mathcal{F}\}$ is linearly independent and hence of size at most $\dim(\mathbb{Q}^n) = n$. This would readily imply $|\mathcal{F}| \leq n$.

To this end, suppose for a contradiction that there exist rationals $\lambda_A : A \in \mathcal{F}$ such that

$$\sum_{A \in \mathcal{F}} \lambda_A \chi_A = 0.$$

Fix some $B \in \mathcal{F}$ and take the standard scalar product of both sides with χ_B. We get

$$\left(\sum_{A \in \mathcal{F}} \lambda_A \chi_A\right) \cdot \chi_B = \sum_{A \in \mathcal{F}} \lambda_A (\chi_A \cdot \chi_B) = \sum_{A \in \mathcal{F}} \lambda_A |A \cap B| = \lambda_B |B| + t \sum_{A \in \mathcal{F} \setminus \{B\}} \lambda_A = 0.$$

Subtracting and adding the term $t\lambda_B$ gives

$$\lambda_B(|B| - t) = -t \sum_{A \in \mathcal{F}} \lambda_A =: -t\Lambda.$$

If $\Lambda \neq 0$, then Λ and λ_B have opposite sign. Since B was arbitrary, this would hold for all $B \in \mathcal{F}$. But this is a contradiction since $\Lambda = \sum_{B \in \mathcal{F}} \lambda_B$.

So, $\Lambda = 0$. It follows that either $\lambda_B = 0$ or $|B| - t = 0$. It cannot be the latter since we assumed $|A| > t$ for all $A \in \mathcal{F}$. Thus, $\lambda_B = 0$. But since B was an arbitrary set in $\mathcal{F}$, this implies all λ_B are zero. Hence, $\{\chi_A \mid A \in \mathcal{F}\}$ are indeed linearly independent.

□

A similar argument can be made in the modular setting.

Theorem 2 (Oddtown) *Let $\mathcal{F} \subseteq 2^{[n]}$. Suppose $|A|$ is odd for all $A \in \mathcal{F}$ and $|A \cap B|$ is even for all $A, B \in \mathcal{F}$. Then $|\mathcal{F}| \leq n$.*

Proof The proof is analogous to that of Fisher's inequality, but working over $\mathbb{F}_2$ instead of $\mathbb{Q}$. Let χ_A be the characteristic vector of A over $\mathbb{F}_2$. We claim that $\{\chi_A \mid A \in \mathcal{F}\}$ is linearly independent. Suppose otherwise: then for a non-empty $\mathcal{G} \subseteq \mathcal{F}$ we have

$$\sum_{A \in \mathcal{G}} \chi_A = 0.$$

Taking the 'standard scalar product'[1] with χ_B for some $B \in \mathcal{G}$ we obtain

$$|B| + \sum_{A \in \mathcal{G} \setminus \{B\}} |A \cap B| = 0 \pmod 2,$$

which a contradiction since $|B|$ is odd, and the remaining terms are even. □

Exercise 1 *Rephrase the above proof combinatorially, i.e. without using the language of linear algebra explicitly.*

This can be generalized as follows.

Theorem 3 (Modular Frankl-Wilson Theorem) *Let p be a prime and $S \subseteq \{0, \ldots, p-1\}$. Let $\mathcal{F} \subseteq 2^{[n]}$ be such that $|A|_{mod\ p} \notin S$ for all $A \in \mathcal{F}$, while $|A \cap B|_{mod\ p} \in S$ for all $A \neq B \in \mathcal{F}$. Then,*

$$|\mathcal{F}| \leq \sum_{i=0}^{|S|} \binom{n}{i}.$$

Proof For each $A \in \mathcal{F}$, let χ_A be the characteristic vector of A over the finite field $\mathbb{F}_p$. Furthermore, for each $A \in \mathcal{F}$, we define the n-variate polynomial $f_A : \mathbb{F}_p^n \to \mathbb{F}_p$ via

$$f_A(x_1, \ldots, x_n) = \prod_{s \in S} \left(\sum_{i \in A} x_i - s \right).$$

Then, for $B \in \mathcal{F}$, $B \neq A$, we have

$$f_A(\chi_B) = \prod_{s \in S} (|A \cap B| - s) = 0.$$

[1] Not a scalar product in the classical sense, but a bilinear form nonetheless.

On the other hand, when $A = B$ we have

$$f_A(\chi_A) = \prod_{s \in S} \left(\sum_{i \in A} |A| - s \right) \neq 0,$$

since $|A|_{\text{mod } p} \notin S$.

Now consider the operation $\sim$ on multivariate polynomials variables which reduces all exponents to 1. For example, $f(x, y, z) = x^2y^3 + z^4 + 3 \xrightarrow{\sim} \tilde{f}(x, y, z) = xy + z + 3$. Notice that for 0-1 vectors $v \in \{0, 1\}^n$ we have $f(v) = \tilde{f}(v)$. So, for $A, B \in \mathcal{F}$ it holds that $\tilde{f}_A(\chi_B) = 0$ if $B \neq A$ and $\tilde{f}_B(\chi_B) \neq 0$. We now claim that the set of polynomials $\{\tilde{f}_A \mid A \in \mathcal{F}\}$ is linearly independent over $\mathbb{F}_p$.

Suppose that there are coefficients $\lambda_A \in \mathbb{F}_p : A \in \mathcal{F}$ such that

$$\sum_{A \in \mathcal{F}} \lambda_A \tilde{f}_A = 0,$$

and let us evaluate this at χ_B for some $B \in \mathcal{F}$. We obtain

$$\sum_{A \in \mathcal{F}} \lambda_A \tilde{f}_A(\chi_B) = \lambda_B \tilde{f}_B(\chi_B) + \sum_{A \in \mathcal{F} \setminus \{B\}} \lambda_A \tilde{f}_A(\chi_B) = \lambda_B \tilde{f}_B(\chi_B) = 0,$$

and since $\tilde{f}_B(\chi_B) \neq 0$ we must have $\lambda_B = 0$. Since B was chosen from $\mathcal{F}$ arbitrarily, all λ_A must be equal zero, meaning the polynomials $\tilde{f}$ are linearly independent, as claimed.

Since each $\tilde{f}_A$ is multi-linear of total degree at most $|S|$, we conclude that

$$|\mathcal{F}| \leq \dim \operatorname{span}\{\tilde{f}_A : A \in \mathcal{F}\} \leq \sum_{i=0}^{|S|} \binom{n}{i}.$$

$\square$

Remark 1 *It is natural to ask whether Frankl-Wilson generalizes to non-prime residues. This was shown to not be true by Grolmusz, who has shown that for $p = 6$, $\mathcal{F}$ can be superpolynomial in size.*